

Poluvodički integrirani sklopovi, minijaturizacija i rakovo oko

I. dio

Zvonimir Ogorelec, Zagreb

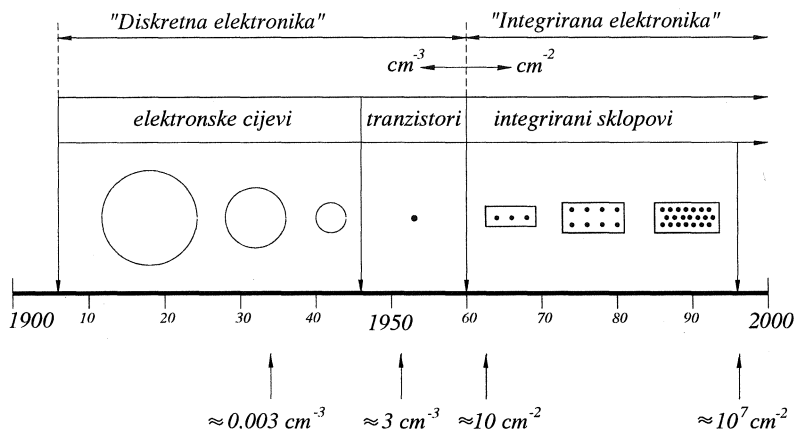
Proslov

Pored niza predavanja, čiji cilj bijaše pokazati kakva je uloga računala u nekim temeljnim istraživanjima u fizici, organizatori ovogodišnje, već trinaeste po redu Ljetne škole mladih fizičara, predvidjeli su i jedno predavanje s temom malo drugačijeg značaja. S temom, naime, koja bi dirnula u pitanje kakva se fizika “krije” u samim računalima. Ili, u drugačije sročeno pitanje, na koji je način i uolikoj mjeri fizika pridonijela dosadašnjem razvoju računala i što se od nje može očekivati u njihovom daljnjem razvoju. To je, naravno, vrlo komplicirana i razgranata tema, jer su za nastanak i usavršavanje računala zaslužna vrlo različita područja fizike. Ipak, bez velike štete i nepravde prema ostalim područjima, značenje fizike se može u dobroj mjeri prikazati i tako da se opiše njena uloga u razvoju bitnih dijelova današnjih računala, a to su njihovi mikroelektronički sklopovi. Budući da je riječ o poluvodičkim sklopovima – a u njihovom se razvoju najjasnije uočava trajna težnja sve većoj složenosti i sve manjim dimenzijama – svoje mjesto u takvom prikazu moraju posve sigurno naći neki aspekti moderne fizike poluvodiča, ali i mnogo širi problem – problem fizičkih ograničenja upravo spomenutoj težnji. Prikaz bi, napokon, bio krajnje siromašan kad se u njemu ne bi našla barem skica nekih sasvim novih ideja. Premda je predavanje, koje je nastojalo udovoljiti ovim zahtjevima, popraćeno tekstom u Zborniku ljetne škole, autor je poslušao sugestije nekih svojih kolega, malo taj tekst preuredio i nadopunio te ga tako priredio za tisak i u ovom časopisu. S ciljem, dakako, da bude nadohvat ruku i širem krugu znatiželjnih čitatelja. Autoru je i ovom zgodom pripomenuti kako su i predavanja i ovaj članak nastali u godini jedne važne obljetnice. To je obljetnica otkrića tranzistora, svakako najvažnije sastavnice današnjih mikroelektroničkih sklopova.

Kratki podsjetnik

Usprkos činjenici da se o tranzistoru, a pogotovo o mikroelektroničkim sklopovima govori danas najviše u okviru elektronike, ona je u ovom tekstu slabo zastupljena. Podsjetnik, međutim, na neke općenite karakteristike elektroničkih sklopova ipak se nameće kao zacijelo najbolji uvod u fiziku koja leži u njihovim temeljima. U najkraćim crtama elektronički se sklop definira kao skup aktivnih i pasivnih komponenti povezanih međusobno u jedinstvenu cjelinu čija struktura ovisi o namjeni sklopa. Aktivne komponente su elektronske cijevi, tranzistori, diode i tome slično, a pasivne otpornici, kapacitori i zavojnice. Sklopovi mogu biti analogni i digitalni. U ovom drugom, za računalne sklopove važnijem slučaju, oni funkcioniraju po zakonitostima Booleove algebre koja operira samo s dvije varijable: 1 i 0. Govori se još i o Booleovoj logici kad se te varijable označavaju s *DA* i *NE*. Uvijek kad se radi o digitalnim sklopovima, aktivne se komponente mogu poistovjetiti s bistabilnim prekidačima. U poluvodičkim mikroelektroničkim sklopovima – koji su dodatak “mikro” dobili zbog svojih malih

dimenzija – ulogu tih prekidača igraju tranzistori. Kakvi, pak, bili da bili – analogni ili digitalni – za sve elektroničke sklopove vrijedi vrlo jednostavna zakonitost: što kompliciraniju zadaću moraju obaviti to im broj komponenti mora biti veći. Tako se jednostavni radioprijamnik može sastaviti uz pomoć tri do četiri komponente, za neki kontrolni uređaj potrebno ih je već stotinjak, a za primitivno računalo barem deset tisuća.



Sl. 1. Minijaturizacija elektroničkih komponenti kroz stotinjak godina.
Krugovi sve manjeg promjera na lijevoj strani slike označavaju konstrukciju sve manjih elektronskih cijevi, a pravokutnici sa sve većim brojem točaka na desnoj strani slike sve veću složenost poluvodičkih integriranih sklopova.

Makar to u njenim ranim danima nije bilo jasno izrečeno, elektronika se oduvijek nastojala prilagoditi jednim te istim zahtjevima. Pokretačka sila njena razvoja, naime, uvijek se mogla razaznati iz ovih nekoliko težnji: iz težnje što manjim dimenzijama aktivnih komponenti (kako bi i kompliciran uređaj, sastavljen od velikog broja komponenti, bio "podnošljiv" dimenzija), zatim iz težnje brzom odzivu aktivnih komponenti (kako bi i čitav uređaj bio brza odziva), iz težnje što manjem pasivnom utrošku energije (zbog ekonomičnosti i mogućnosti gradnje pokretnih uređaja) i, naposljetku, iz težnje što većoj pouzdanosti (kako bi se komponente, a time i čitav sklop što manje kvarili). Pobrojani zahtjevi nisu ni izdaleka neovisni, ali se najgrublje rečeno ipak svode samo na prvi – na što manje dimenzije. Elektronika je, dakle, već od svojih ranih dana težila minijaturizaciji sklopova. Iz sl. 1, gdje je ukratko i na slikovit način prikazan njezin tijek kroz povijest elektronike, vidljivo je da se težnja ispočetka iskazivala konstrukcijom sve manjih i manjih ("minijaturnih") elektronskih cijevi. Zatim je neposredno nakon drugog svjetskog rata, 1947. godine, izbila "prva elektronička revolucija", kad je minijaturizacija dramatično uznapredovala otkrićem tranzistora, no tu je "revoluciju" ubrzo, već koncem pedesetih godina, zamijenila "druga elektronička revolucija", i to idejom o gradnji monolitno integriranih poluvodičkih sklopova. Njihovo usavršavanje, dakako, traje još i danas. Ta druga "revolucija", uzgred, uvela je i dva nova pojma. Elektroniku do nje ljudi obično nazivaju "diskretnom" elektronikom. Ovu, pak, noviju – "integriranom" elektronikom. Kao mjera uspješnosti minijaturizacije obično se navodi gustoća slaganja elektroničkih komponenti. Ona je od 1935. godine (kad su se koristile "velike" elektronske cijevi) do 1955. godine (kad su već uvelike zavladali tranzistori) porasla od oko 0.003 do oko 3 komponente po kubičnom centimetru – punih, eto, tri reda veličine. Za uvid u daljnji napredak, prostorna gustoća nije više pogodna veličina, jer se monolitno integrirani sklopovi karakteriziraju površinskom gustoćom slaganja. A ona je od početka šezdesetih godina do danas narasla od nekoliko desetaka do nekoliko milijuna komponenti po kvadratnom centimetru silicijske pločice.

Tema ovogodišnje Ljetne škole i njena ostala predavanja sugerirali su još jednu, nesumnjivo uvjerljivu ilustraciju brzog razvoja elektronike. To je ilustracija usporedbom negdašnjih i današnjih računala. Poredba je doista impresivna, makar se ponekad zlorabi i koristi s natruhama omalovažavanja prijašnjih znanstvenih i tehnoloških dostignuća. Riječ je ponajprije o opisu jednog od prvih digitalnih računala, glasovitog ENIAC-a. To računalo (Electronic Numerical Integrator And Computer) izgrađeno je 1946. godine na Moore School of Electrical Engineering, na američkom Pennsylvanijskom sveučilištu. Njegove aktivne komponente bile su elektronske cijevi, ukupno njih 18 000. Logičke sklopove računala konstruktori su raspodijelili u 40 memorijskih i procesorskih modula s ukupnom masom od oko tridesetak tisuća kilograma. Čitav uređaj bio je smješten u dvije prostorije s površinom kakvu imaju prosječne gimnastičke vježbaonice. Dimenzijama golemo računalo bilo je golemo i po utrošku električne energije. Budući da žarna nit jedne elektronke zahtijeva snagu od oko 3 W, samo za grijanje cijevi računala valjalo je rezervirati snagu od oko 50 kW. Uzevši drugačije, žarne niti svih tih cijevi predstavljale su moćnu električnu peć, pa je zbog stvaranja podnošljive radne atmosfere obje prostorije trebalo hladiti industrijskim ventilatorima. Pouzdanost uređaja, međutim, nije bila bogzna kakva. Premda je elektronska cijev, sama za se, srazmjerno pouzdan pribor, vjerojatnost kvara kod tako velikog broja cijevi bila je ipak velika. Zaista, u vrijeme uhodavanja uređaja pregorijevale su prosječno dvije cijevi dnevno. Brigu oko njihove zamjene morao je voditi samo za to imenovan i stalno dežuran tehničar. Na kraju, kao važnu karakteristiku svakog računala treba spomenuti i radnu frekvenciju ENIAC-a. Ona je iznosila skromnih stotinjak kHz, barem za tri reda veličine manje od takta danas već široko rasprostranjenih osobnih računala. Ipak, bez obzira na goleme dimenzije, na rastrošnost i na skromne karakteristike, u razmišljanju o ENIAC-u nikad se ne smije javiti omalovažavanje. To računalo je bilo i trajno će ostati značajno znanstveno, tehnološko, u krajnjoj crti – civilizacijsko dostignuće.

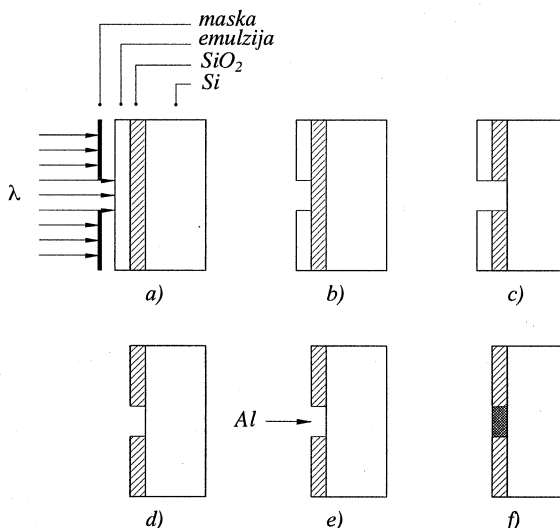
Na žalost, kao i mnoga druga materijalna dostignuća, ENIAC nije imao sretnu sudbinu. Nakon skoro jednog desetljeća uspješnog rada “rashodovali” su ga 1955. godine. Potom su ga bez ikakve potrebe rastavili na dijelove, a njegove module raspršili po brojnim američkim tehničkim muzejima. Na Moore School of Electrical Engineering, mjestu rođenja ENIAC-a, preostala su još svega četiri komada. No, u predvorju gdje te relikvije stoje na ogled, nalazi se i jedva vidljiv “čip”. A do njega natpis: *“U manje od četrdeset godina poluvodička je tehnologija toliko uznapredovala da može proizvesti računala superiornija ENIAC-u, a čije se komponente mogu smjestiti na komadić silicija sa stranicama od svega četvrtine inča”*.

Fizička ograničenja

U rečenom predvorju, eto, možda se najbolje vidi kako se dramatično i u relativno kratkom vremenu izmijenila bit elektronike. Fizika čvrstog stanja, posebice fizika poluvodiča, sve šire i dublje znanje o strukturi tih materijala, o utjecaju defekata na njihova svojstva i, uopće, sve jasnija kvantnomehanička predodžba o njihovom elektronskom sustavu, zaista je pripomogla izbijanju niza već spomenutih elektroničkih “revolucija”. No, bez obzira na ondašnje ili današnje stanje, nešto je ipak ostalo isto. Razvoj elektronike, sada već mikroelektronike, i dalje pokreću ona ista četiri zahtjeva koja su njen razvoj poticala i u prošlosti. Ljudi se, dakle, i dalje trse graditi što manje, a kompleksnije uređaje te i dalje nastoje da im odziv bude što brži, a da pritom ne trpe na pouzdanosti i štedljivosti. Dosadašnji komentar tih zahtjeva čini ih dosta razumljivima, no iscrpno odgovoriti na pitanje zašto je tome baš tako nije ipak jednostavno. Riječ je o vrlo složenom skupu zahtjeva u kojem se isprepliću mnoge, čak i suprotne stvari. Osnovni dojmovi, dakako, mogu se steći i svojevrsnim prećicama. Po jednoj od njih

takav se dojam stječe kratkom analizom međuovisnosti veličine uređaja i njegova odziva. Uzmimo da je riječ o elektroničkom uređaju koji mora obaviti neku vrlo kompliciranu zadaću. To po već spomenutom pravilu znači da mora imati golem broj komponenti. Sve te komponente i pomoćni pribor zauzimaju neki volumen koji karakterizira linearna dimenzija L . Pod uvjetom da se električni signal širi kroz sklop svojom maksimalnom brzinom, a to je brzina svjetlosti c , njemu će za put L biti potrebno vrijeme $t = L/c$. Zahtijeva li konstrukcija uređaja takozvani sinkroni rad, što znači da se signal u različitim dijelovima sklopa smije razlikovati pomakom faze manjim od "clock perioda", onda L uređaja određuje i minimalni "clock period". Kao što se lijepo vidi, "brži" uređaj (viša radna frekvencija) mora imati manje dimenzije (manji L). Na primjer, želi li se da period iznosi 4 ns (kao u jednom od starijih superračunala Cray), karakteristična dimenzija L uređaja ne smije biti veća od 1.2 m. Ako se sve potrebne komponente ne mogu smjestiti u taj volumen, uređaj željenog odziva ne može se ni izgraditi.

Naravno, račun je toliko pojednostavljen da ga je teško usporediti sa stvarnošću. U prvom redu, brzina rada uređaja može se povećati mimo ovog zahtjeva drugačijom arhitekturom sklopa, njegovim rastavljanjem na podsustave i takozvanim asinkronim radom. To u ovom tekstu nije zanimljivo, jer se suština fizičkih zakonitosti uopće ne mijenja. Čitav se problem samo spušta s razine sustava na razinu podsustava. Ono što je kudikamo zanimljivije jest činjenica da je opisani zahtjev čak i stroži, uzme li se u obzir da brzina rada digitalnog sklopa ovisi i o tome kako se brzo prekopčavaju njegovi prekidači, to jest, kojom brzinom prelaze tranzistori iz vodljivog u nevodljivo stanje i obratno. Budući da vrijeme prekopčavanja ovisi i opet o dimenzijama (o debljini baze bipolarnog ili o dužini upravljačke elektrode unipolarnog tranzistora), zahtjev za daljnjom minijaturizacijom sklopa dobiva još više na snazi. S pravom se postavlja pitanje što ta nastojanja ograničava u fizičkom smislu.



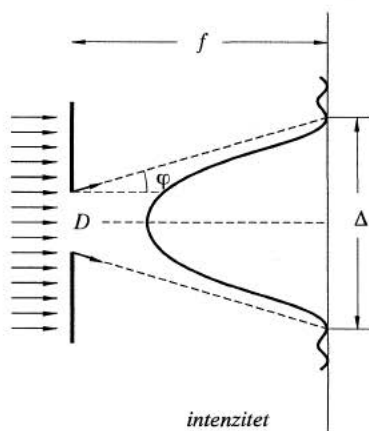
Sl. 2. Skica sukcesivnih postupaka planarne tehnologije pri stvaranju vodljive linije:

- a) snimanje pukotine na fotoosjetljivu emulziju, b) otapanje osvijetljene emulzije,
- c) jetkanje sloja SiO₂, d) uklanjanje neosvijetljene emulzije, e) naparavanje metala,
- f) gotova linija.

Problem je najlakše raščlaniti tako da se prati stvaranje najjednostavnije "komponente" poluvodičkog integriranog sklopa, a to je vodič koji spaja dvije aktivne komponente

– “vodljiva linija”, kako je obično nazivaju. U prvim monolitnim sklopovima njena je širina iznosila stotinjak mikrometara da bi se do danas smanjila na ispod jednog mikrometra. To je bez sumnje velik uspjeh, no za njegovo razumijevanje potrebno je barem djelomično poznavanje takozvane planarne tehnologije kojom se današnji “čipovi” proizvode. Riječ je, u stvari, o vrlo kompliciranom procesu koji se sastoji čak od petnaest-dvadeset sukcesivnih potprocesa. U svakom, pak, od njih javlja se jedan ili više ovakvih postupaka: oksidiranje silicijske pločice, prekrivanje oksida tankim slojem posebne fotoosjetljive emulzije (takozvanog fotorezista), pravljenje “maske”, to jest, fotografiranje i smanjivanje crteža budućeg sklopa ili njegovih dijelova, preslikavanje “maske” na fotoosjetljivu emulziju (takozvana fotolitografija), odstranjivanje neosvijetljenih dijelova emulzije, jetkanje nepokrivenog silicij dioksida, difuziju ili implantaciju primjesnih atoma, naparavanje metalnih vodiča i kontakata, spajanje dovodnih žica i još mnogo toga. Proces je skiciran na sl. 2.

Na sreću, unatoč istinskoj složenosti tog tehnološkog postupka, osnovni dojam o njegovim ograničenjima može se steći na temelju jednostavne predodžbe. Budući da se stvaranje vodljive linije u tom procesu zasniva na fotografiranju, na “crtanju” kontura sklopa uz pomoć svjetlosnog snopa, problem širine vodljive linije može se sa stanovišta fizike načeti pitanjem koliko se uska linija uopće može nacrtati takvim postupkom. Bitno ograničenje krije se u pojavi ogiba svjetlosti. Opišimo ga na ovom jednostavnom pokusu (sl. 3).



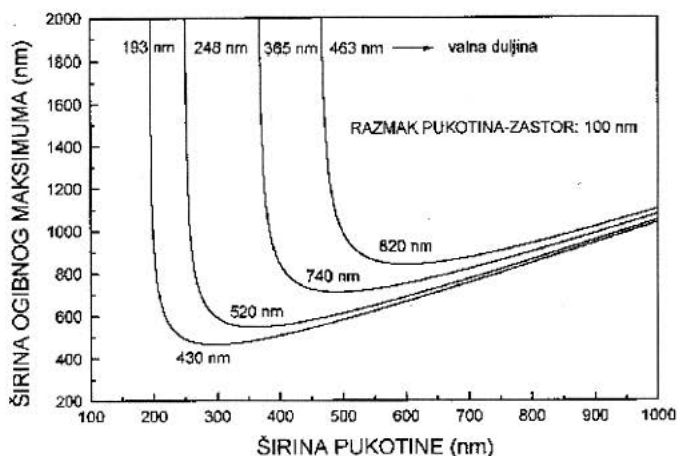
Sl. 3. Prikaz ogiba svjetlosti u Fraunhoferovoj aproksimaciji.

Neka na neprozirnu foliju s pukotinom širine D (to je kontura vodljive linije na “maski” sklopa) pada paralelni snop svjetlosti valne duljine λ . Na zastoru, koji je za f udaljen od pukotine (to je udaljenost “maske” od površine oksida, što se u “kontaktnoj fotolitografiji” svodi na debljinu emulzije), javlja se slika pukotine. Ona, naravno, nije geometrijska kopija pukotine, već njena ogibna slika. Lijevo i desno od središta slike javljaju se izmjenično jače i slabije osvijetljena mjesta. U središnjem ogibnom maksimumu skoncentrirano je najviše svjetlosnog toka (više od 80%), pa se tok svjetlosti u ostalim maksimumima može zanemariti. Najvažnija od svega jest činjenica da je širina Δ središnjeg maksimuma uvijek veća od širine pukotine D . To u tehnološkom smislu znači da je minimalna širina linije koja se može doći uvijek veća od širine pukotine koja služi kao “maska”. Ilustrirajmo tu zakonitost brojkama.

Teorija ogiba u Fraunhoferovoj aproksimaciji daje za širinu prvog ogibnog maksimuma ovu relaciju:

$$\Delta = D + \frac{2f\lambda}{\sqrt{D^2 - \lambda^2}}.$$

Iz nje se odmah vidi da će Δ uz konstantnu D biti to manja što je valna duljina λ upotrebene svjetlosti manja. Ne uvijek, naravno. Jer, ako se λ približava vrijednosti D , drugi član na desnoj strani izraza sve više i više raste te u graničnom slučaju $\lambda = D$ postaje beskonačno velik. Zaključak je, dakle, jasan: tehnologija teži stvaranju sve užih pukotina i fotolitografiji sa svjetlošću sve kraćih valnih duljina. Na primjer, uzmemo li za udaljenost f realnu vrijednost 100 nm i pretpostavimo li da se koristi svjetlost ovih valnih duljina: 436 nm, 365 nm, 248 nm i 193 nm (to je ljubičasto i ultraljubičasto zračenje posebnih lasera udomaćenih u današnjoj tehnologiji), Fraunhoferova relacija daje krivulje prikazane na sl. 4. Iz nje se vidi da odgovarajuće minimalne širine linija iznose redom: 820 nm, 740 nm, 520 nm i 430 nm.



Sl. 4. Ovisnost širine prvog ogibnog maksimuma o širini pukotine za svjetlost četiriju valnih duljina uobičajenih u planarnoj tehnologiji.

Desna strana slike sugerira sve manji utjecaj ogiba, a lijeva sve brži i neograničeni porast širine linija. Iako je netom spomenutih 430 nm uistinu mala vrijednost, ne treba je uzeti kao konačnu. Postoji još nekoliko "trikova" (smanjivanje f , primjerice) koji vodljivu liniju mogu još više suziti. Prognoze zato govore da će do kraja stoljeća, to jest tisućljeća širina vodljive linije pasti na ispod 200 nm. Pritom je nužno razlikovati rezultate postignute u proizvodnji i laboratorijske rezultate. Na pokusnim uzorcima rezultati su redovito bolji. Budući da je ipak riječ o približavanju već vrlo strogim fizičkim ograničenjima, uže linije sve se teže i teže dostižu. Izlaz je očito u nekoj, kažimo radikalnoj tehnološkoj promjeni. Jedna od njih je prijelaz na fotolitografiju sa zračenjem dramatično kraćih valnih duljina, na primjer, rendenskog zračenja ($\lambda = 0.4 - 5$ nm). Istini za volju, ideja nije nova, ali s njenim ostvarenjem nitko nije bio zadovoljan. Tehnologija se, doduše, tome nastojala prilagoditi. Umjesto običnih "maski" izmišljene su "maske" od tankih slojeva za zračenje nepropusnih materijala i tek su na njih "nacrtane" konture budućeg sklopa. Na žalost, divergentno zračenje ostavlja iza "maske" nedopustivo široke i nejednolike "sjene", što kvari potencijalno dobre rezultate. No, čudnim slučajem rendenska je fotolitografija ipak postala vrlo privlačnim rješenjem.